sfCrime.r

Gaurab

Fri Sep 12 21:55:16 2014

library(ggplot2)  
crDf = read.csv("sfCrime.csv", header=TRUE, fill=TRUE)  
# Exploratory Data Analysis  
summary(crDf)

## IncidntNum Category   
## Mode:logical LARCENY/THEFT :8205   
## NA's:30760 OTHER OFFENSES:4004   
## NON-CRIMINAL :3653   
## ASSAULT :2518   
## VEHICLE THEFT :1885   
## VANDALISM :1611   
## (Other) :8884   
## Descript DayOfWeek   
## GRAND THEFT FROM LOCKED AUTO : 3166 Friday :4583   
## STOLEN AUTOMOBILE : 1307 Monday :4415   
## DRIVERS LICENSE, SUSPENDED OR REVOKED: 1304 Saturday :4394   
## AIDED CASE, MENTAL DISTURBED : 1142 Sunday :4357   
## LOST PROPERTY : 1038 Thursday :4303   
## PETTY THEFT OF PROPERTY : 955 Tuesday :4294   
## (Other) :21848 Wednesday:4414   
## Date Time PdDistrict   
## 06/28/2014: 452 18:00 : 681 SOUTHERN :6185   
## 06/29/2014: 432 12:00 : 650 MISSION :4011   
## 08/09/2014: 422 19:00 : 643 CENTRAL :3867   
## 06/27/2014: 411 20:00 : 627 NORTHERN :3205   
## 08/08/2014: 408 17:00 : 611 BAYVIEW :2970   
## 06/20/2014: 406 22:00 : 559 INGLESIDE:2613   
## (Other) :28229 (Other):26989 (Other) :7909   
## Resolution Address   
## NONE :19527 800.0 Block of BRYANT ST : 960   
## ARREST, BOOKED : 6237 800.0 Block of MARKET ST : 324   
## ARREST, CITED : 2559 900.0 Block of POTRERO AV : 219   
## PSYCHOPATHIC CASE: 1144 1000.0 Block of POTRERO AV: 203   
## LOCATED : 807 2000.0 Block of MISSION ST: 149   
## UNFOUNDED : 149 16TH ST / MISSION ST : 133   
## (Other) : 337 (Other) :28772   
## X Y   
## Min. :-123 Min. :37.7   
## 1st Qu.:-122 1st Qu.:37.8   
## Median :-122 Median :37.8   
## Mean :-122 Mean :37.8   
## 3rd Qu.:-122 3rd Qu.:37.8   
## Max. :-122 Max. :37.8   
##   
## Location   
## (37.7752316978411, -122.403742962696): 851   
## (37.7568815847358, -122.406662657485): 213   
## (37.7568256145719, -122.406656623063): 198   
## (37.7650501214965, -122.419671780296): 178   
## (37.7847532835996, -122.407036790381): 157   
## (37.7751775060495, -122.403675173997): 101   
## (Other) :29062

head(crDf)

## IncidntNum Category Descript DayOfWeek  
## 1 NA LARCENY/THEFT GRAND THEFT FROM UNLOCKED AUTO Sunday  
## 2 NA LARCENY/THEFT GRAND THEFT FROM LOCKED AUTO Sunday  
## 3 NA LARCENY/THEFT GRAND THEFT FROM LOCKED AUTO Sunday  
## 4 NA DRUG/NARCOTIC POSSESSION OF METH-AMPHETAMINE Sunday  
## 5 NA DRUG/NARCOTIC POSSESSION OF COCAINE Sunday  
## 6 NA LARCENY/THEFT GRAND THEFT FROM LOCKED AUTO Sunday  
## Date Time PdDistrict Resolution Address  
## 1 08/31/2014 20:30 CENTRAL NONE HYDE ST / CALIFORNIA ST  
## 2 08/31/2014 14:30 CENTRAL NONE COLUMBUS AV / JACKSON ST  
## 3 08/31/2014 11:30 CENTRAL NONE SUTTER ST / STOCKTON ST  
## 4 08/31/2014 17:49 MISSION ARREST, BOOKED 16TH ST / MISSION ST  
## 5 08/31/2014 18:05 NORTHERN ARREST, BOOKED LARKIN ST / OFARRELL ST  
## 6 08/31/2014 00:01 CENTRAL NONE GRANT AV / SUTTER ST  
## X Y Location  
## 1 -122.4 37.79 (37.7909741243888, -122.417392830334)  
## 2 -122.4 37.80 (37.7963018736036, -122.404417620748)  
## 3 -122.4 37.79 (37.7894347630337, -122.406958660602)  
## 4 -122.4 37.77 (37.7650501214965, -122.419671780296)  
## 5 -122.4 37.79 (37.7851670875814, -122.417903977564)  
## 6 -122.4 37.79 (37.7896302267231, -122.405402610955)

names(crDf)

## [1] "IncidntNum" "Category" "Descript" "DayOfWeek" "Date"   
## [6] "Time" "PdDistrict" "Resolution" "Address" "X"   
## [11] "Y" "Location"

nrow(crDf)

## [1] 30760

# Where shouldn't you park your car?  
uniDescription = crDf[!duplicated(crDf["Descript"]),"Descript"]  
  
# From the uniDescription, we get a sense of the description which tells us where we should not park cars.   
# We find the words AUTO and VEHICLE which indicates crimes related to cars. We should target these places not to park cars.  
  
toMatch = c("AUTO", "VEHICLE")  
add = crDf[grep(paste(toMatch, collapse="|"), crDf[,"Descript"]),c("Address")]  
df = as.data.frame(table(add))  
df$rank = rank(-df$Freq,ties.method="min")  
df = df[order(df$rank,decreasing = F),]  
  
# Top 10 addresses where most of the AUTO/VEHICLE related thefts have happened  
df[1:10,]

## add Freq rank  
## 5314 800.0 Block of BRYANT ST 303 1  
## 4379 400.0 Block of STOCKTON ST 47 2  
## 4686 500.0 Block of JOHNFKENNEDY DR 42 3  
## 2439 200.0 Block of 3RD ST 37 4  
## 3330 2600.0 Block of GEARY BL 33 5  
## 4028 3600.0 Block of LYON ST 32 6  
## 2585 200.0 Block of INTERSTATE80 HY 27 7  
## 1189 10TH ST / HARRISON ST 22 8  
## 1155 1000.0 Block of POINTLOBOS AV 21 9  
## 3908 3200.0 Block of 20TH AV 21 9

# Bar plot showing the same  
q = qplot(x=add, y=Freq, data=df[1:10,], geom="bar", stat="identity")  
q + theme(axis.text.x = element\_text(angle = 90, hjust = 1))

